Taylor & Francis
Taylor & Francis Group

Journal of Quality Technology
A Quarterly Journal of Methods, Applications and Related Topics

ISSN: 0022-4065 (Print) 2575-6230 (Online) Journal homepage: https://www.tandfonline.com/loi/ujqt20

Multivariate Control Charts for Individual Observations

Nola D. Tracy, John C. Young & Robert L. Mason

To cite this article: Nola D. Tracy, John C. Young & Robert L. Mason (1992) Multivariate Control Charts for Individual Observations, Journal of Quality Technology, 24:2, 88-95, DOI: 10.1080/00224065.1992.12015232

To link to this article: https://doi.org/10.1080/00224065.1992.12015232

Published online: 11 May 2018.

Submit your article to this journal

Article views: 3

View related articles

Citing articles: 52 View citing articles

Full Terms & Conditions of access and use can be found at
https://www.tandfonline.com/action/journalInformation?journalCode=ujqt20

Multivariate Control Charts for Individual Observations

NOLA D. TRACY, JOHN C. YOUNG

McNeese State University, Lake Charles, LA 70609-2340

ROBERT L. MASON

Southwest Research Institute, San Antonio, TX 78228-0510

When p correlated process characteristics are being measured simultaneously, often individual observations are initially collected. The process data are monitored and special causes of variation are identified in order to establish control and to obtain a "clean" reference sample to use as a basis in determining the control limits for future observations. One common method of constructing multivariate control charts is based on Hotelling's T^2 statistic. Currently, when a process is in the start-up stage and only individual observations are available, approximate F and chi-square distributions are used to construct the necessary multivariate control limits. These approximations are conservative in this situation. This article presents an exact method, based on the beta distribution, for constructing multivariate control limits at the start-up stage. An example from the chemical industry illustrates that this procedure is an improvement over the approximate techniques, especially when the number of subgroups is small.

Introduction

THE quality of the output of a production process is often measured by the joint level of several correlated characteristics. For example, a chemical process may be a function of temperature and pressure, both of which need to be monitored carefully; a particular grade of lumber might depend on correlated characteristics such as stiffness and bending strength. Hawkins (1974) refers to a geochemical process in coal mining in which each observation consists of 14 correlated characteristics. In these types of situations, separate univariate control charts for each characteristic are often utilized to detect changes in the inherent variability of the process. When these characteristics are mutually correlated, however, the univariate charts are not as sensitive as multivariate methods that capitalize on the correlation. While Hotelling (1947) was among the first to note the drawbacks in using separate univariate control charts for correlated variables, many current quality control books (e.g., Ryan (1989)) contain examples of multivariate processes where the use of separate individual charts would not have detected out-of-control conditions. Many of the concepts of multivariate quality control are due to Hotelling (1947). Excellent summaries and discussions of these techniques are found in Alt (1985), Jackson (1980, 1981a, 1981b, 1985, 1991), Ryan (1989), and Montgomery (1991). The values plotted on multivariate control charts are usually statistics based on his well-known T^2 distribution (Hotelling (1931)). This distribution is the multivariate counterpart to Student's t distribution. The multivariate T^2 chart is particularly appropriate when the characteristics of interest are correlated.

There are two distinct phases in constructing control charts (e.g., see Alt (1982)). The first phase, which offers a retrospective view, involves testing whether the process was in control when the initial individual or subgroup data were collected on the process. A subgroup represents a sample of observations taken at some point in the process, such as a sample taken during a specified time period. This phase is often

Dr. Tracy is an Assistant Professor in the Department of Mathematics, Computer Science, and Statistics.

Dr. Young is a Professor in the Department of Mathematics, Computer Science, and Statistics.

Dr. Mason is Manager of the Statistical Analysis Section. He is a Senior Member of ASQC.

termed the start-up stage of the process as the purpose is to obtain a set of data (a reference sample) to establish the control limits for monitoring purposes. The goal of the first stage is to establish statistical control (i.e., a "clean" process) and find accurate control limits for stage two. The second phase consists of using the control chart to maintain control, that is, detecting any departure from the process standards as future subgroups are drawn.

The multivariate T^2 statistic often is utilized as the charting statistic for both phases of control chart construction. In the first phase, with subgroup sizes greater than one, and in the second phase, where concern is in maintaining control of a process, the control limits are determined using the fact that the T^2 statistic (times a constant) follows an exact F distribution. In the start-up stage with individual data, however, the control limits are computed using an approximation based on the F or chi-square distributions. The degree of error associated with the approximation is unknown. The purpose of this paper is to address this problem by presenting an exact method for constructing a multivariate control chart for use when individual observations are collected in the start-up stage of the process. The results are illustrated using an actual data example taken from the chemical industry.

Establishing Control in the Start-Up Stage

Consider the case where p correlated characteristics are being measured simultaneously and are in need of control. Assume that these characteristics follow a p-dimensional multivariate normal distribution with mean vector $\boldsymbol{\mu}' = (\mu_1, \mu_2, \ldots, \mu_p)$ and covariance matrix Σ, where μ_i is the mean for the i^{th} characteristic and Σ is a $p \times p$ matrix consisting of the variances and covariances of the p characteristics. The multivariate normal distribution is the p-dimensional analogue to the univariate normal distribution assumed for each characteristic. Note that technically if a process is not in statistical control, as in the start-up situation assumed here, then there is no stable distribution for the data. The assumption of multivariate normality is being made here solely for the purpose of deriving the control limits. After control has been established we are assuming that the data are reasonably normally distributed. Our results depend on the validity of this assumption, just as the validity of the usual control limits for a univariate control chart for individuals requires the assumption of normality. The assumption of multivariate normality can be checked using appropriate multivariate normal goodness-of-fit tests (e.g., see Gnanadesikan (1977)).

Assume that control of the process is in the start-up stage and a sample of m subgroups of past data are available to estimate the parameters $\boldsymbol{\mu}$ and Σ. In some situations it may not be possible to take subgroups of size larger than one. This can occur either when the production rate is too slow to conveniently allow subgroup sizes greater than one or when repeated measurements differ only because of laboratory or analysis error, as in many chemical processes.

For notational purposes, represent the i^{th} individual observation of the p characteristics from the reference sample with the vector

$$\mathbf{X}_i = \begin{bmatrix} X_{i1} \\ X_{i2} \\ \vdots \\ X_{ip} \end{bmatrix}.$$

The estimated mean vector, whose components are the means of each characteristic, is

$$\bar{\mathbf{X}}_i = \begin{bmatrix} \bar{X}_1 \\ \bar{X}_2 \\ \vdots \\ \bar{X}_p \end{bmatrix}$$

where

$$\bar{\mathbf{X}}_j = \frac{1}{m} \sum_{i=1}^{m} \mathbf{X}_{ij}$$

and the estimated covariance matrix is

$$\mathbf{S}_m = \frac{1}{m-1} \sum_{i=1}^{m} (\mathbf{X}_i - \bar{\mathbf{X}}_m)(\mathbf{X}_i - \bar{\mathbf{X}}_m)'.$$

To construct a multivariate control chart based on Hotelling's T^2 statistic, for observation $\mathbf{X}_i$ one uses the charting statistic

$$Q_i = (\mathbf{X}_i - \bar{\mathbf{X}}_m)'\mathbf{S}_m^{-1}(\mathbf{X}_i - \bar{\mathbf{X}}_m). \qquad (1)$$

The distribution of Q_i is not widely known, and thus most multivariate control chart practitioners approximate it (see examples in Jackson (1985) or Ryan (1989)) with a chi-square distribution or an F distribution to obtain the control chart limits.

If one assumes that the estimates of $\bar{\mathbf{X}}_m$ and $\mathbf{S}_m$ are the true population values $\boldsymbol{\mu}$ and Σ, respectively, then Seber (1984, pp. 18-19) has shown that the statistic Q_i is distributed as a chi-square variate with p degrees of freedom. In that case, the lower control limit is

$$LCL = \chi^2(1 - \alpha/2;\ p) \qquad (2)$$

and the upper control limit is

$$UCL = \chi^2(\alpha/2; p) \tag{3}$$

where $\chi^2(\alpha; p)$ is the $1 - \alpha$ percentile of the chi-distribution with p degrees of freedom.

If one assumes that the i^{th} observation $\mathbf{X}_i$ is independent of both $\overline{\mathbf{X}}_m$ and $\mathbf{S}_m$, then the statistic Q_i (times a constant) follows an F distribution with p and $m - p$ degrees of freedom (see the Appendix for details). In that case the lower control limit is

$$LCL = \frac{p(m-1)(m+1)}{m(m-p)} F(1 - \alpha/2; p, m - p)$$

and the upper control limit is

$$UCL = \frac{p(m-1)(m+1)}{m(m-p)} F(\alpha/2; p, m - p)$$

where $F(\alpha; p, m - p)$ is the $1 - \alpha$ percentile of the F distribution with p and $m - p$ degrees of freedom.

Since neither of the above assumptions holds true in the start-up stage described here, the suggested approximations for the distribution of the charting statistic Q_i have some drawbacks. For example, unless p is small, a large sample is required for the chi-square distribution to fit (see Hawkins (1981)). Fortunately, these problems may be avoided since it is possible to derive the exact distribution of Q_i. Gnanadesikan and Kettenring (1972), based on a result of Wilks (1962), have shown that Q_i (times a constant) has a beta distribution. Specifically,

$$Q_i \sim \frac{(m-1)^2}{m} B(p/2, (m - p - 1)/2). \tag{4}$$

The distribution in (4) is correct only when individual $\mathbf{X}_i$ values collected in the start-up stage of the process (i.e., used to compute the control limits) are checked to see whether they fall within the control limits. In contrast, when future observations are taken on the "clean process" and checked against the control limits calculated from the start-up data, the statistics that are formed are independent of $\overline{\mathbf{X}}_m$ and $\mathbf{S}_m$ and follow an exact F distribution. This will be discussed in more detail later.

Knowing the correct distribution of Q_i, it is possible to construct the needed control limits. The lower control limit is given by

$$LCL = \frac{(m-1)^2}{m} B(1 - \alpha/2; p/2, (m - p - 1)/2)$$

and the upper control limit is given by

$$UCL = \frac{(m-1)^2}{m} B(\alpha/2; p/2, (m - p - 1)/2)$$

where $B(\alpha; p/2, (m - p - 1)/2)$ is the $1 - \alpha$ percentile of the beta distribution with parameters $p/2$ and $(m - p - 1)/2$. If tables for the beta distribution are not readily available, the relationship

$$\frac{(p/(m-p-1))F(\alpha; p, m-p-1)}{1 + (p/(m-p-1))F(\alpha; p, m-p-1)} = B(\alpha; p/2, (m - p - 1)/2)$$

between random variables with beta and F distributions can be utilized. Applying this relationship gives the limits

$$LCL = \frac{(m-1)^2}{m} \times \frac{(p/(m-p-1))F(1 - \alpha/2; p, m-p-1)}{1 + (p/(m-p-1))F(1 - \alpha/2; p, m-p-1)} \tag{5}$$

and

$$UCL = \frac{(m-1)^2}{m} \times \frac{(p/(m-p-1))F(\alpha/2; p, m-p-1)}{1 + (p/(m-p-1))F(\alpha/2; p, m-p-1)} \tag{6}$$

in terms of percentiles from the F distribution.

In many situations the LCL is set to zero. The reason for this is that any shift in the mean will lead to an increase in the statistic Q_i, and thus the LCL can be ignored. However, Q_i is sensitive not only to shifts in the mean vector but also to changes in the covariance matrix of the data. If the covariance matrix were to change, it might result in abnormally small values of Q_i. Hence, to detect such changes we have chosen to use a nonzero LCL. It should be noted that large values of Q_i can also be caused by changes in the covariance matrix and not just by changes in the mean vector (see Hawkins (1991)).

Center lines are commonly displayed on univariate control charts. This practice would be especially helpful on multivariate control charts since the upper and lower limits are not spaced symmetrically about the median. A reasonable center line for this type of multivariate chart can be obtained using equation (6) with $\alpha = 1$ (i.e., using the 50^{th} percentile of the F distribution).

Example

Consider the data set given in Table 1, which represents actual data taken in a start-up stage of a chemical industrial process. The application and the data have been disguised to protect proprietary information. Like many chemical processes, this example involves the simultaneous measurement of three variables: percentage of impurities (X_1), temperature (X_2), and concentration strength (X_3) of a particular substance. Preliminary tests provided no reason to doubt that the data follow a multivariate normal distribution.

In this sample, there are 14 observations on three variables, so $m = 14$ and $p = 3$. The vector of sample means is

$$\bar{\mathbf{X}}_{14} = \begin{bmatrix} 16.83 \\ 85.19 \\ 43.21 \end{bmatrix}$$

and the sample covariance matrix is

$$\mathbf{S}_{14} = \begin{bmatrix} 0.365 & -0.022 & 0.100 \\ -0.022 & 1.036 & -0.245 \\ 0.100 & -0.245 & 0.224 \end{bmatrix}.$$

The sample correlation matrix $\mathbf{R}_m$ is made up of elements r_{ij} representing the pairwise correlation coefficient between $\mathbf{X}_i$ and $\mathbf{X}_j$; that is, the element in the i^{th} row and the j^{th} column of $\mathbf{R}_m$ is given by

$$r_{ij} = \frac{s_{ij}}{\sqrt{s_{ii}}\sqrt{s_{jj}}}$$

where s_{ij} is the element in the i^{th} row and the j^{th} column of the sample covariance matrix $\mathbf{S}_m$. For this example, the correlation matrix $\mathbf{R}_{14}$ is

$$\mathbf{R}_{14} = \begin{bmatrix} 1.000 & -0.035 & 0.348 \\ -0.035 & 1.000 & -0.507 \\ 0.348 & -0.507 & 1.000 \end{bmatrix}.$$

Although the collected data are from the start-up stage and do not necessarily represent an in-control process, we can see from the off-diagonal elements of the correlation matrix that at this point the three variables are pairwise correlated. Thus, a multivariate control chart is appropriate.

Control limits and a charting statistic are needed to construct a multivariate control chart. Using equation (1), the values of the charting statistic Q_i are given in Table 1. The corresponding control limits using equations (5) and (6) with $\alpha = 0.01$ are

$$LCL = \frac{(14-1)^2}{14} \times \frac{(3/(14-3-1)) \times 0.0229}{1 + (3/(14-3-1)) \times 0.0229} = 0.082$$

$$UCL = \frac{(14-1)^2}{14} \times \frac{(3/(14-3-1)) \times 8.081}{1 + (3/(14-3-1)) \times 8.081} = 8.55.$$

Figure 1 shows the corresponding multivariate control chart. Observations 1 and 5 both lie outside of the control limits. These observations were examined individually to determine a possible assignable cause. It was determined that the unusually low level of im-

TABLE 1. Chemical Industry Data and Statistics

Obs. #	% Impurities	Temp	Concentration	Q_i ($m = 14$)	Q_i ($m = 13$)
1	14.92	85.77	42.26	10.93	
2	16.90	83.77	43.44	2.04	1.84
3	17.38	84.46	42.74	5.58	5.33
4	16.90	86.27	43.60	3.86	3.58
5	16.92	85.23	43.18	0.04	0.23
6	16.71	83.81	43.72	2.25	2.17
7	17.07	86.08	43.33	1.44	1.46
8	16.93	85.85	43.41	1.21	1.05
9	16.71	85.73	43.28	0.68	1.91
10	16.88	86.27	42.59	2.17	5.16
11	16.73	83.46	44.00	4.17	3.84
12	17.07	85.81	42.78	1.40	1.65
13	17.60	85.92	43.11	2.33	7.00
14	16.90	84.23	43.48	0.90	0.77

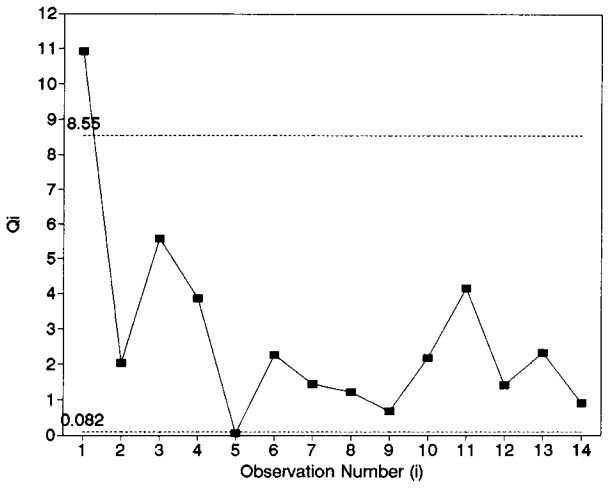

FIGURE 1. Multivariate Control Chart for Original Data Set with 14 Observations.

purities for observation 1 was the result of a sampling error. Therefore, this point was removed from the sample. Since no assignable cause could be associated with observation 5, it was retained in the sample.

Examining the observations individually works in this application. In general, however, interpreting a multivariate control chart does not simply reduce to an examination of the univariate charts for the original variables. It follows from the definition of Q_i in equation (1) that points outside the multivariate control limits are a result of one or more principal components being out of control. Proper interpretation requires consideration of these principal components (see Jackson (1991)).

Removing observation 1 and recalculating the parameter estimates with $m = 13$, a new vector of estimated means

$$\bar{\mathbf{X}}_{13} = \begin{bmatrix} 16.98 \\ 85.14 \\ 43.28 \end{bmatrix}$$

as well as a new estimated covariance matrix

$$\mathbf{S}_{13} = \begin{bmatrix} 0.068 & 0.076 & -0.055 \\ 0.076 & 1.092 & -0.216 \\ -0.055 & -0.216 & 0.163 \end{bmatrix}$$

and a new sample correlation matrix

$$\mathbf{R}_{13} = \begin{bmatrix} 1.000 & 0.280 & -0.520 \\ 0.280 & 1.000 & -0.512 \\ -0.520 & -0.512 & 1.000 \end{bmatrix}$$

are obtained. In $\mathbf{R}_{13}$ the three variables remain correlated but the correlations between $\mathbf{X}_1$ and $\mathbf{X}_2$ and between $\mathbf{X}_1$ and $\mathbf{X}_3$ are reversed in signs. This is a direct result of the removal of observation 1, which had an unusually low reading on $\mathbf{X}_1$, the percent of impurities.

The recalculated Q_i values are given in the last column of Table 1. The corresponding control limits for this sample (now of size 13) are

$$LCL = \frac{(13-1)^2}{13} \times \frac{(3/(13-3-1)) \times 0.0228}{1 + (3/(13-3-1)) \times 0.0228} = 0.084$$

$$UCL = \frac{(13-1)^2}{13} \times \frac{(3/(13-3-1)) \times 8.717}{1 + (3/(13-3-1)) \times 8.717} = 8.24.$$

Replotting these Q_i values, we observe in Figure 2 that none of the observations are outside the control limits. Statistical control has been established by eliminating the special cause of variation, namely the aberrant point caused by sampling error. We now have a reference sample that can be used in calculating the control limits for the second stage. It is interesting to note that observation 5, which was below the LCL with the full sample ($m = 14$), is now inside the control limits. The dependence among the Q_i's contributes to this.

Maintaining Control With Future Values

The second phase in the construction of multivariate control charts consists of testing to see whether

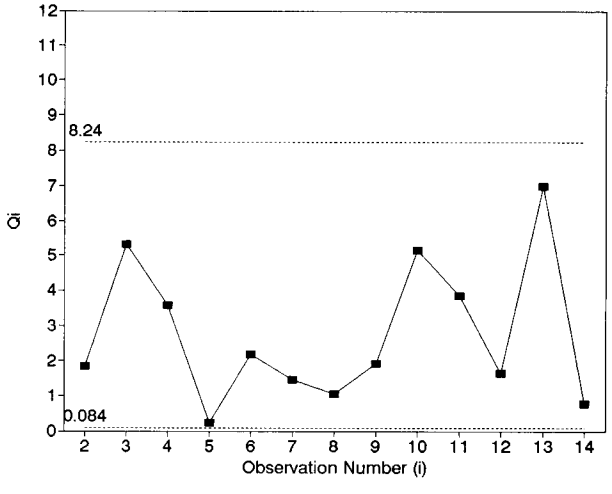

FIGURE 2. Multivariate Control Chart for Modified Data Set with 13 Observations.

the process remains in control as future subgroups are drawn. At this stage the vector of means $\bar{\mathbf{X}}_m$ and the covariance matrix $\mathbf{S}_m$ obtained in the start-up stage are used to calculate control limits, which will be used to test the future observations. Thus, a future observation $\mathbf{X}_f$ is independent of $\bar{\mathbf{X}}_m$ and $\mathbf{S}_m$. We use Hotelling's statistic

$$T_f^2 = (\mathbf{X}_f - \bar{\mathbf{X}}_m)'\mathbf{S}_m^{-1}(\mathbf{X}_f - \bar{\mathbf{X}}_m)$$

where $\mathbf{X}_f$ denotes the p-dimensional vector of future observations on the p characteristics, $\bar{\mathbf{X}}_m$ is the p-dimensional mean vector of the m observations in the "clean" reference sample, and $\mathbf{S}_m$ is the $p \times p$ covariance matrix associated with these observations.

If the start-up sample size is large, a common approach is to assume the estimates $\bar{\mathbf{X}}_m$ and $\mathbf{S}_m$ from the start-up phase are "standards" and equal the true population parameters $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$. The statistic T_f^2 would then have the form

$$T_f^2 = (\mathbf{X}_f - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{X}_f - \boldsymbol{\mu})$$

and would follow a chi-square distribution with p degrees of freedom. The resulting upper and lower multivariate control limits are the same as those given in equations (2) and (3). However, these are approximations since $\bar{\mathbf{X}}_m$ and $\mathbf{S}_m$ are not population parameters but random variables.

There is no need for the above chi-square approximation since the exact distribution of T_f^2 can be obtained (shown in the Appendix). Its distribution is

$$T_f^2 \sim \frac{p(m+1)(m-1)}{m(m-p)} F(p, m-p).$$

Thus, the exact control limits are

$$LCL = \frac{p(m+1)(m-1)}{m(m-p)} F(1-\alpha/2;\, p, m-p)$$

$$UCL = \frac{p(m+1)(m-1)}{m(m-p)} F(\alpha/2;\, p, m-p).$$

For the data in Table 1 the control limits would be

$$LCL = \frac{3(13+1)(13-1)}{13(13-3)} \times 0.0229 = 0.088$$

$$UCL = \frac{3(13+1)(13-1)}{13(13-3)} \times 8.081 = 31.33.$$

For the future vector of observations

$$\mathbf{X}_f = \begin{bmatrix} 17.08 \\ 84.08 \\ 43.81 \end{bmatrix}$$

$T_f^2 = 3.52$. Since this value falls well within the control limits, the process at this observation would be in control.

Note that, using the exact method, the UCL for maintaining control with future observations is roughly four times larger than the UCL obtained in establishing control in the start-up phase. Thus, treating the start-up sample mean vector and covariance matrix as if they were independent of the observations leads to a very conservative UCL when there is a small number of subgroups. In contrast, the chi-square approximation in equation (3) would yield $UCL = 12.84$ for both situations. This would be conservative for the start-up data but liberal for future data.

Comparison of Exact and Approximate Methods

A simple comparison of the exact and approximate distributions suggested for establishing control in the start-up phase of the process can be made by comparing the values of the upper control limits. Table 2 gives the upper control limit for a multivariate control chart using the exact distribution (beta) and the approximations (F and chi-square) with $\alpha = 0.01$ and $p = 2, 5, 10$. Note that for a small number of subgroups, such as $m = 20$, with $p = 5$ the F-approximated UCL is 35.72, the chi-square-approximated UCL is 16.7, and the exact UCL is 12.01. With $p = 10$ the F approximation is $UCL = 116.64$, the chi-square approximation is $UCL = 25.19$, and the exact UCL is 15.83. This illustrates the conservative error that results from the use of approximate UCL's in the start-up phase. As the number of variables increases, the difference between the approximate UCL's becomes even more evident. One can easily extend these comparisons for any value of m, p, or α using tables of the F, chi-square, and beta distributions.

If the approximate methods discussed earlier for the start-up stage had been employed in our data example, the calculated control limits with the full data set would have been different. Use of the F distribution yields an approximate $LCL = 0.082$ and an approximate $UCL = 29.65$. While the LCL is exactly the same, the UCL is much larger than the exact value of 8.55. Use of the chi-square approximation results in a $LCL = 0.072$ and an $UCL = 12.84$. Again, both approximations yield conservative estimates of the UCL.

Conclusion

Analysis using multivariate T^2 control charts can be a powerful tool in process control situations involving simultaneous measurements of several char-

TABLE 2. Multivariate Upper Control Limits for $\alpha = 0.01$ and $p = 2, 5, 10$

	$p = 2$				$p = 5$				$p = 10$		
m	Beta	F	Chi-square	m	Beta	F	Chi-square	m	Beta	F	Chi-square
1			10.60	1			16.75	1			25.19
2			10.60	2			16.75	2			25.19
3		79998.00	10.60	3			16.75	3			25.19
4	2.25	597.00	10.60	4			16.75	4			25.19
5	3.18	132.80	10.60	5			16.75	5			25.19
6	4.04	65.71	10.60	6		576394.95	16.75	6			25.19
7	4.78	43.95	10.60	7	5.14	2989.50	16.75	7			25.19
8	5.39	33.94	10.60	8	6.11	529.60	16.75	8			25.19
9	5.90	28.35	10.60	9	7.02	224.60	16.75	9			25.19
10	6.32	24.85	10.60	10	7.82	134.50	16.75	10			25.19
11	6.67	22.46	10.60	11	8.52	95.53	16.75	11		2422448.70	25.19
12	6.98	20.74	10.60	12	9.13	74.82	16.75	12	10.08	10966.90	25.19
13	7.24	19.44	10.60	13	9.66	62.26	16.75	13	11.07	1747.43	25.19
14	7.46	18.44	10.60	14	10.12	53.96	16.75	14	11.99	681.40	25.19
15	7.66	17.63	10.60	15	10.53	48.12	16.75	15	12.82	381.30	25.19
20	8.37	15.23	10.60	20	12.01	34.02	16.75	20	15.83	111.09	25.19
25	8.81	14.05	10.60	25	12.92	28.57	16.75	25	17.67	70.78	25.19
30	9.10	13.34	10.60	30	13.54	25.71	16.75	30	18.90	55.78	25.19
50	9.69	12.10	10.60	50	14.81	21.28	16.75	50	21.39	38.18	25.19
70	9.95	11.64	10.60	70	15.36	19.78	16.75	70	22.47	33.40	25.19
100	10.14	11.30	10.60	100	15.77	18.77	16.75	100	23.28	30.47	25.19

acteristics. A representative reference sample is essential; it is important to base the constructed control limits on accurate estimates of the parameters. During the start-up stage, when using subgroups consisting of individual observations (i.e., subgroups of size 1) with measurement variables, the beta distribution should be utilized to obtain control limits for the T^2 statistic (i.e., the Q_i's). Use of this exact distribution is better than employing approximate F and chi-square distributions, especially when the number of subgroups is small, a condition apt to be prevalent in start-up situations.

Acknowledgments

The authors wish to thank Dr. John Cornell, Dr. Peter Nelson, and the referees for their many helpful comments and suggestions which greatly improved this paper. Dr. Tracy was supported in part by a grant from the Shearman Research Initiative Fund.

Appendix

Theorem (Seber (1984, pp. 30–31)): Let $T^2 = m\mathbf{y}'\mathbf{W}^{-1}\mathbf{y}$, where $\mathbf{y} \sim N_d(\mathbf{0}, \Sigma)$, $\mathbf{W} \sim W_d(m, \Sigma)$, and $\mathbf{y}$ and $\mathbf{W}$ are statistically independent. N_d and W_d are used to denote d-dimensional normal and Wishart distributions, respectively. Then

$$\frac{m - d + 1}{d}\frac{T^2}{m} \sim F(d, m - d + 1).$$

Using the notation of this paper, consider a set of initial multivariate observations $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_m$, and a future observation $\mathbf{X}_f$, where each $\mathbf{X}_i$ is a vector of observations on p variables. If

$$\mathbf{X}_i \sim N_p(\boldsymbol{\mu}, \Sigma)$$

then

$$\bar{\mathbf{X}}_m \sim N_p(\boldsymbol{\mu}, \Sigma/m)$$

$$(m - 1)\mathbf{S}_m \sim \mathbf{W}_p(m - 1, \Sigma).$$

Now suppose that $\mathbf{X}_f$, $\bar{\mathbf{X}}_m$, and $\mathbf{S}_m$ are independent, as is the case when $\bar{\mathbf{X}}_m$ and $\mathbf{S}_m$ are calculated from the start-up data and $\mathbf{X}_f$ is a future observation. Then

$$\mathbf{X}_f - \bar{\mathbf{X}}_m \sim N_p\left(\mathbf{0}, \left(\frac{m + 1}{m}\right)\Sigma\right)$$

and

$$\sqrt{\frac{m}{m + 1}}(\mathbf{X}_f - \bar{\mathbf{X}}_m) \sim N_p(\mathbf{0}, \Sigma).$$

If one defines the statistic

$$T^2 = \left(\frac{m}{m+1}\right)(\mathbf{X}_f - \overline{\mathbf{X}}_m)'\mathbf{S}_m^{-1}(\mathbf{X}_f - \overline{\mathbf{X}}_m)$$

then

$$\frac{(m-p)}{p}\frac{T^2}{(m-1)} \sim F(p, m-1-p+1)$$

which leads to

$$\frac{m-p}{p(m-1)}\frac{m}{(m+1)}(\mathbf{X}_f - \overline{\mathbf{X}}_m)'\mathbf{S}_m^{-1}(\mathbf{X}_f - \overline{\mathbf{X}}_m) \sim F(p, m-p)$$

and

$$(\mathbf{X}_f - \overline{\mathbf{X}}_m)'\,\mathbf{S}_m^{-1}(\mathbf{X}_f - \overline{\mathbf{X}}_m) \sim \frac{p(m-1)(m+1)}{m(m-p)}F(p, m-p).$$

References

ALT, F. B. (1982). "Multivariate Quality Control: State of the Art". *ASQC Quality Congress Transactions*, pp. 886–893.

ALT, F. B. (1985). "Multivariate Quality Control". *Encyclopedia of Statistical Sciences* 6 (S. Kotz and N. Johnson, eds.), John Wiley & Sons, New York, NY, pp. 110–122.

GNANADESIKAN, R. and KETTENRING, J. R. (1972). "Robust Estimates, Residuals, and Outlier Detection with Multiresponse Data". *Biometrics* 28, pp. 81–124.

GNANADESIKAN, R. (1977). *Methods for Statistical Data Analysis of Multivariate Observations.* John Wiley & Sons, New York, NY.

HAWKINS, D. M. (1974). "The Detection of Errors in Multivariate Data Using Principal Components". *Journal of the American Statistical Association* 69, pp. 340–344.

HAWKINS, D. M. (1981). "A New Test for Multivariate Normality and Homoscedasticity". *Technometrics* 23, pp. 37–44.

HAWKINS, D. M. (1991). "Multivariate Quality Control Based on Regression-Adjusted Variables". *Technometrics* 33, pp. 61–76.

HOTELLING, H. (1931). "The Generalization of Student's Ratio". *Annals of Mathematical Statistics* 2, pp. 360–378.

HOTELLING, H. (1947). "Multivariate Quality Control". *Techniques of Statistical Analysis* (C. Eisenhart, M. Hastay, and W. A. Wallis, eds.), McGraw-Hill, New York, NY, pp. 111–184.

JACKSON, J. E. (1980). "Principal Components and Factor Analysis: Part I—Principal Components". *Journal of Quality Technology* 12, pp. 201–213.

JACKSON, J. E. (1981a). "Principal Components and Factor Analysis: Part II—Additional Topics Related to Principal Components". *Journal of Quality Technology* 13, pp. 46–58.

JACKSON, J. E. (1981b). "Principal Components and Factor Analysis: Part III—What is Factor Analysis?". *Journal of Quality Technology* 13, pp 125–130.

JACKSON, J. E. (1985). "Multivariate Quality Control". *Communications in Statistics* 14 (11), pp. 2657–2688.

JACKSON, J. E. (1991). *A User's Guide to Principal Components.* John Wiley & Sons. New York, NY.

MONTGOMERY, D. C. (1991). *Introduction to Statistical Quality Control,* 2nd ed. John Wiley & Sons, New York, NY.

RYAN, T. P. (1989). *Statistical Methods for Quality Improvement.* John Wiley & Sons, New York, NY.

SEBER, G. A. F. (1984). *Multivariate Observations.* John Wiley & Sons, New York, NY.

WILKS, S. S. (1962). *Mathematical Statistics.* John Wiley & Sons, New York, NY.

Key Words: *Multivariate Quality Control, T^2 Control Charts.*